## Contents

<table>
<thead>
<tr>
<th>Acknowledgements</th>
<th>ix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abbreviations</td>
<td>xi</td>
</tr>
</tbody>
</table>

1 What is Computer Corpus Lexicography (CCL)?  
1.0 Introduction  
1.1 Developing notions of the lexicon  
1.1.1 The trend towards lexicalism  
1.1.2 The word vs the clause as the most basic unit in linguistics  
1.2 Lexical content  
1.2.1 Characterisations of lexical content  
1.2.2 Hudson's All-Inclusive Lexicon  
1.2.3 The TEI Lexicon  
1.3 Lexical form  
1.4 Conclusion and the organisation of this book  
1.5 Study questions  
1.6 Further reading  
   Notes  
2 The lexicon in Computational Linguistics (CL1), Computational Lexicography (CL2), and Corpus Linguistics (CL3)  
2.0 Introduction  
2.1 What is Computational Linguistics (CL1)?  
   2.1.1 The lexicon in CL1  
2.2 What is Computational Lexicography (CL2)?  
   2.2.1 Human consumption vs machine-use  
   2.2.2 The lexicon in CL2  
2.3 What is Corpus Linguistics (CL3)?  
   2.3.1 The lexicon in CL3  
2.4 The interdependence and convergence of CL1, CL2, and CL3  

2.5 The TEI for the emergence of textual and lexical standards 39
2.6 The notion of the reusability of lexical resources 40
2.7 The core-reusable multifunctional lexicon 42
2.8 The polytheoretical lexicon 43
2.9 Conclusion 44
2.10 Study questions 45
2.11 Further reading
Notes 45

3 Corpus evidence and lexicon-based language modelling 47
3.0 Introduction 47
3.1 Types of lexicographic and lexical evidence 47
3.2 The corpus as a lexical resource: the issue of representativeness 52
3.3 Corpus size and the 'Bank of English' 55
3.4 The search for units of meaning 57
3.5 Conclusion 63
3.6 Study questions 63
3.7 Further reading
Notes 65

4 Methods of lexical acquisition: 'Learning' a lexicon for real texts from real text 67
4.0 Introduction 67
4.1 The relation between the lexicon and the corpus 67
4.2 Manual lexical acquisition 73
4.3 Automatic and semi-automatic lexical acquisition 75
4.4 The lexicographer/linguist's workbench for lexical acquisition 80
4.5 A framework for lexical analysis 86
4.6 Conclusion 87
4.7 Study questions 88
4.8 Further reading
Notes 89

5 Computational storage of the lexicon 91
5.0 Introduction 91
5.1 The question of formalism 91
5.2 Lexical database vs lexical knowledge base? 94
5.3 The notion of inheritance 95
5.4 The DATR lexical knowledge representation language 97
5.5 The ACQUILEX lexical knowledge base 100
5.6 The PROLEX Prolog lexical database 101
5.7 The WORDNET lexical database of English 104
5.8 The CELEX lexical database 106
5.8.1 The CELEX English lexical database 107
5.8.2 The structure of the CELEX English database 108
5.8.3 Accessing the database 109
5.8.4 Selected lexemes using the CELEX linguistic codes 109
5.8.5 General comments on the CELEX database 118
5.9 Using the World Wide Web for lexical research 118
5.10 Conclusion 119
5.11 Study questions 119
5.12 Further reading 119
Notes 121

6 A case study: Applying the LFA framework to two corpora of business English 123
6.0 Introduction 123
6.1 The notions of sublanguage, genre, and register 123
6.2 The PROLEX corpus 125
6.2.1 The PROLEX lexicon 129
6.3 The PROCOMPARE corpus 131
6.3.1 The PROCOMPARE lexicon 134
6.4 Corpus tagging 135
6.5 Corpus syntactic parsing 137
6.6 Corpus word-extraction 143
6.7 Corpus collocation 143
6.8 Corpus semantic tagging and parsing 145
6.9 Corpus pragmatics 145
6.9.1 The COCOA format for referencing the PROLEX corpus 148
6.9.2 Extraction of typological frequency from the PROLEX corpus 148
6.10 Structuring lexical entries from the PROLEX and PROCOMPARE corpora 149
6.10.1 Morphology 151
6.10.2 Syntax 152
6.10.3 Semantics 157
6.10.4 Pragmatics 158
6.11 A discussion of the PROLEX and PROCOMPARE specimen lexical entries 161
6.11.1 Account (n) 161
6.11.2 Advise (v) 162
6.11.3 Appreciate (v) 163
6.11.4 Claim (v, n) 163
6.11.5 Full (adj) 164
6.11.6 Please (adv) 165
6.11.7 Regret (v) 165
6.11.8  Shipment (n) 165
6.11.9  The relatedness of information 166
6.11.10 Making the LFA lexicon more compact 167
6.12  Conclusion 171
6.13  Study questions 172
6.14  Further reading 172
Notes 172

7  Conclusion: Towards a more adequate concept of the lexicon 173

Appendix A: Specimen lexical entries 177
Appendix B: Some World Wide Web sites for CL1, CL2, and CL3 211
Appendix C: Suggested solutions to exercises 215
Bibliography 219
Index 241